

A Comprehensive Survey of Graph Embedding: Problems, Techniques and Applications

Hongyun Cai, Vincent W. Zheng, and Kevin Chen-Chuan Chang

- graph embedding aims to represent a graph as low dimensional vectors while the graph structures are preserved.
 - graph analytics aims to mine useful information from graph data
 - representation learning obtains data representations that make it easier to extract useful information when building classifiers or other predictors

简介

- 图分析
 - node classification
 - node clustering
 - node retrieval/recommendation
 - link prediction
- 图种类
 - homogeneous graph
 - heterogeneous graph
 - graph with auxiliary information
 - graph constructed from non-relational data.
- 不同粒度

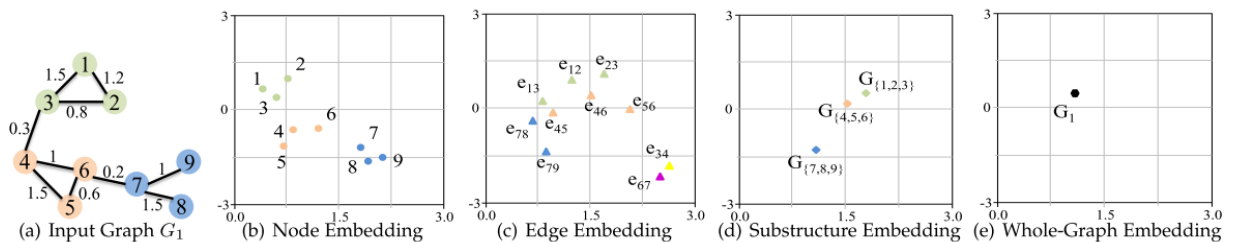


Fig. 1. A toy example of embedding a graph into 2D space with different granularities. $G_{\{1,2,3\}}$ denotes the substructure containing node v_1, v_2, v_3 .

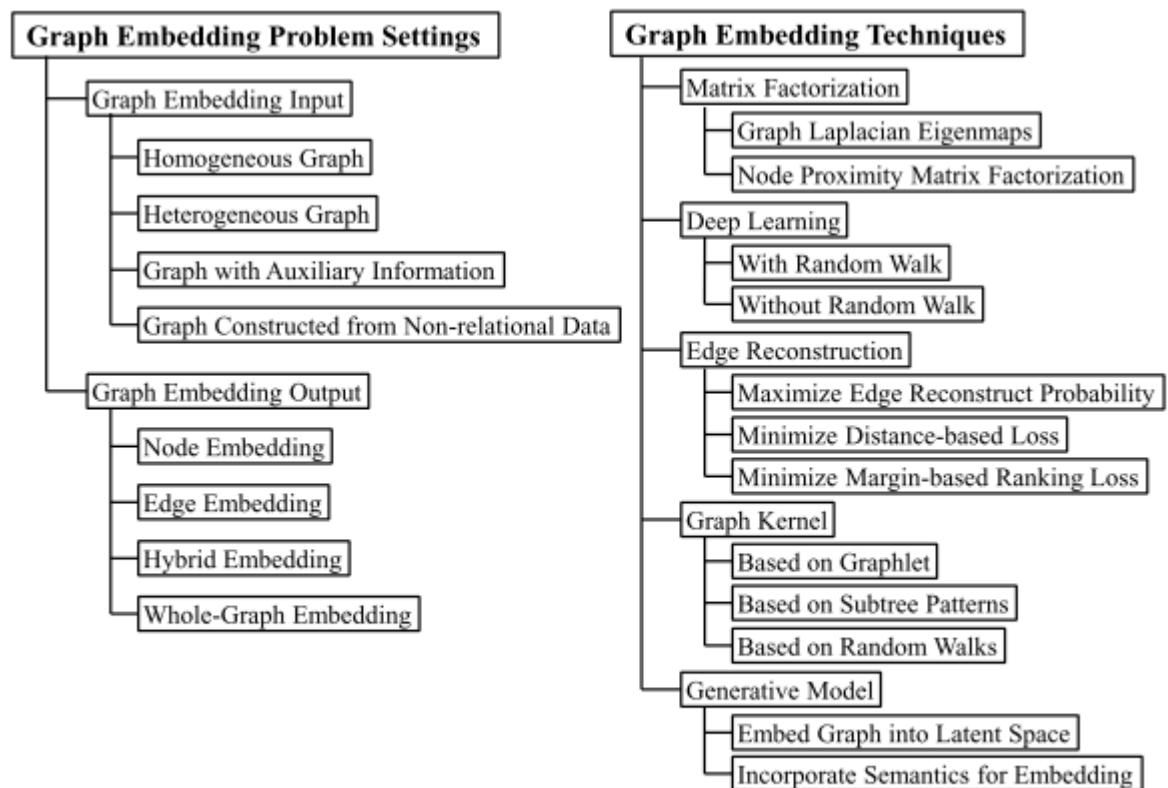


Fig. 2. Graph embedding taxonomies by problems and techniques.

挑战

Graph Embedding Input

- Homogeneous Graph
 - directed
 - weighted
 - 挑战: How to capture the diversity of connectivity patterns observed in graphs?
- Heterogeneous Graph
 - Community-based Question Answering (cQA) sites
 - Multimedia Networks
 - Knowledge Graphs.
 - 挑战 How to explore global consistency between different types of objects, and how to deal with the imbalances of objects belonging to different types, if any?
- Graph with Auxiliary Information
 - Label
 - Attribute
 - Node feature
 - information propagation
 - Knowledge base
 - 挑战: How to incorporate the rich and unstructured information so that the learnt embeddings are both representing the topological structure and discriminative in terms of the auxiliary information?

- Graph Constructed from Non-relational Data
 - 挑战: How to construct a graph that encodes the pairwise relations between instances and how to preserve the generated node proximity matrix in the embedded space?

Graph Embedding Output

- Node Embedding
- Edge Embedding
- Hybrid Embedding

Graph Embedding Techniques

- Matrix Factorization
 - Graph Laplacian Eigenmaps
 - Node Proximity Matrix Factorization
 - 矩阵分解主要用于非关系型数据中的节点嵌入
- Deep Learning
 - DL based Graph Embedding with Random Walk
 - DL based Graph Embedding without Random Walk
- Edge Reconstruction based Optimization
 - Maximizing Edge Reconstruction Probability
 - Minimizing Distance-based Loss
 - Minimizing Margin-based Ranking Loss
- Graph Kernel
- Generative Model
 - EmbedGraph Into The Latent Semantic Space
 - Incorporate Latent Semantics for Graph Embedding